Structural basis for LMO2-driven recruitment of the SCL: E47bHLH heterodimer to hematopoietic-specific transcriptional targets

Article (Published Version)

Cell Reports
Article

Structural Basis for LMO2-Driven Recruitment of the SCL:E47$_{bHLH}$ Heterodimer to Hematopoietic-Specific Transcriptional Targets

Kamel El Omari,[1] Sarah J. Hoosdally,[2,3] Kapil Tuladhar,[2,3] Dimple Karia,[1,2] Elisa Hall-Ponselé,[2] Olga Platonova,[1] Paresh Vyas,[2] Roger Patient,[2] Catherine Porcher,[2,*] and Erika J. Mancini[1,*]
[1]Division of Structural Biology, The Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford OX3 7BN, UK
[2]MRC Molecular Hematology Unit, Weatherall Institute of Molecular Medicine, University of Oxford, Oxford OX3 9DS, UK
[3]These authors contributed equally to this work
*Correspondence: catherine.porcher@imm.ox.ac.uk (C.P.), erika@strubi.ox.ac.uk (E.J.M.)
http://dx.doi.org/10.1016/j.celrep.2013.06.008

SUMMARY

Cell fate is governed by combinatorial actions of transcriptional regulators assembling into multiprotein complexes. However, the molecular details of how these complexes form are poorly understood. One such complex, which contains the basic-helix-loop-helix heterodimer SCL:E47 and bridging proteins LMO2:LDB1, critically regulates hematopoiesis and induces T cell leukemia. Here, we report the crystal structure of (SCL:E47)$_{bHLH}$:LMO2:LDB1$_{LID}$ bound to DNA, providing a molecular account of the network of interactions assembling this complex. This reveals an unexpected role for LMO2. Upon binding to SCL, LMO2 induces new hydrogen bonds in SCL:E47, thereby strengthening heterodimer formation. This imposes a rotation movement onto E47 that weakens the heterodimer:DNA interaction, shifting the main DNA-binding activity onto additional protein partners. Along with biochemical analyses, this illustrates, at an atomic level, how hematopoietic-specific SCL sequesters ubiquitous E47 and associated cofactors and supports SCL's reported DNA-binding-independent functions. Importantly, this work will drive the design of small molecules inhibiting leukemogenic processes.

INTRODUCTION

Many, if not all, cellular processes are driven by multiprotein complexes. In the nucleus, control of gene expression is achieved through assembly of such molecular machines on genomic loci in a highly controlled manner. These complexes contain combinations of regulators such as transcription factors (TFs), cofactors, and chromatin-remodeling proteins (Goodrich and Tjian, 2010; Malik and Roeder, 2010) that assemble in a modular manner for rapid adaptation of gene expression programs. Although at the heart of transcription, little is known about the molecular mechanisms driving formation of multiprotein complexes and their interaction with DNA, the role of individual components in this process, and the relationship between TFs and chromatin-remodeling proteins. To address some of these questions, we have analyzed the associations between two evolutionarily conserved families of transcriptional regulators, namely basic-helix-loop-helix (bHLH) and LIM domain-containing proteins (German et al., 1992; Kong et al., 1997; Lee and Pfaff, 2003) at structural, functional, and biochemical levels.

The hematopoietic system offers a well-characterized model to study bHLH/LIM protein interactions, such as those engaged by the tissue-specific class II bHLH TF SCL/TAL1 (hereafter called SCL) and the non-DNA-binding LIM-only protein LMO2. SCL and LMO2 were initially discovered through chromosomal translocations involved in T cell acute lymphoblastic leukemia (T-ALL) (Begley et al., 1989; Boehm et al., 1991). Indeed, their ectopic expression in T cell precursors occurs in up to 60% of childhood T-ALL cases of which 80% coexpresses SCL and LMO2 (or LMO1) (Ferrando et al., 2002, 2004). In normal hematopoiesis, SCL and LMO2 are absolutely required for hematopoietic specification and terminal differentiation of specific hematopoietic lineages (Hall et al., 2003; Kassouf et al., 2008; Lécuyer and Hoang, 2004; Patterson et al., 2007; Porcher et al., 1996).

SCL forms obligate heterodimers through its HLH domain with ubiquitously expressed class I bHLH E proteins (such as the E2A gene products, E47 and E12) (Porcher et al., 1999). The SCL:E47 heterodimer binds through its basic regions to an E box DNA recognition sequence (CANNTG), each monomer recognizing one-half of the E box (Massari and Murre, 2000). The SCL:E47 heterodimer then nucleates a "core" multiprotein complex by binding to the adaptor protein LMO2 and its interacting partner LDB1 (LIM-binding domain 1) (Lécuyer and Hoang, 2004).

The SCL core complex (SCL:E47:LMO2:LDB1) acquires further specificity through recruitment of additional protein partners. In erythroid cells, LMO2 recruits hematopoietic-specific

Table 1. Data Collection and Refinement Statistics

Data Collection Details	$(SCL:E47)_{bHLH}$:LMO2:$LDB1_{LID}$:DNA	$(SCL:E47)_{bHLH}$:DNA
Data collection site	I04 Diamond	I02 Diamond
Space group	*F*222	$C222_1$
Cell dimensions (Å, °)	*a* = 102.9, *b* = 141.0, *c* = 148.8	*a* = 73.2, *b* = 152.8, *c* = 55.4
	α = β = γ = 90	α = β = γ = 90
Wavelength (Å)	1.2829	0.9795
Resolution range (Å)	50.0–2.8 (2.9–2.8)	50.0–2.9 (3.0–2.9)
Redundancy	3.8 (2.8)	3.8 (3.7)
Completeness (%)	97.8 (95.5)	97.8 (98.5)
Average I/σ (I)	11.0 (1.8)	15.6 (1.5)
R_{merge}	0.086 (0.511)	0.079 (0.972)
Refinement Statistics		
Resolution range (Å)	30.15–2.80 (2.87–2.80)	42.44–2.87 (2.95–2.87)
R factor (R_{work}/R_{free})	0.224/0.269	0.254/0.288
Rmsd bond length (Å)	0.006	0.005
Rmsd bond angle (°)	0.9	0.8
Mean B factor ($Å^2$)	91.7	117.1
Ramachandran plot (%) favored/allowed/outliers	100.0/95.5/0.0	100.0/99.3/0.0

TF GATA1, creating the so-called ''pentameric'' complex that binds a bipartite E box-GATA DNA sequence (Wadman et al., 1997; Kassouf et al., 2010). The complex regulates expression of important hematopoietic-specific genes (Kassouf et al., 2010) upon recruitment of various combinations of cofactors and chromatin remodelers, such as ETO2, mSin3A, P300, PCAF, and LSD1 (Li et al., 2012; Schuh et al., 2005).

In T-ALL, the prevailing model suggests that ectopically expressed SCL and LMO2 synergize to prevent the activity of E protein homodimers, essential for normal progression of T cell differentiation. Specifically, through a sequestration mechanism that is yet to be characterized at the molecular level, E proteins are locked into ectopic SCL core complexes that act by repressing apoptotic pathways and preventing the normal T cell transcriptional program (Herblot et al., 2000; Palii et al., 2011; Sanda et al., 2012).

Interestingly, for its functions in hematopoietic specification and leukemogenesis, SCL does not rely on direct DNA-binding activities (Draheim et al., 2011; Kassouf et al., 2008), suggesting that it may work off DNA or be tethered to DNA through other DNA-bound TFs. Additionally, we previously showed that one-fifth of SCL's genomic targets in erythroid cells can recruit the factor independently of its DNA-binding activity (Kassouf et al., 2010), suggesting a critical network of additional protein:protein and protein:DNA interactions for the nucleation of such complexes.

Despite a wealth of studies defining the SCL core complex as a key transcriptional regulator in normal and malignant hematopoiesis, its mechanism of action remains undefined at a molecular level. Specifically, the molecular interactions governing protein:protein and protein:DNA associations are not understood. Moreover, the molecular details of the E protein sequestration model, which confers to ectopically expressed SCL and LMO2 their oncogenic properties, are unclear. Although functional, biochemical, and biophysical studies have suggested potential pathways to complex assembly (Lécuyer et al., 2007; Ryan et al., 2008; Schlaeger et al., 2004; Wadman et al., 1997), complete dissection of the molecular rules driving complex formation has been hindered by the lack of a full structural characterization. No structure exists for SCL, and whereas other bHLH proteins and LIM proteins, including E47 (Ellenberger et al., 1994) and LMO2 (El Omari et al., 2011), have been structurally characterized, no molecular account exists of their association. The mechanistic role of each complex component and their contribution to cofactor recruitment also remain unclear.

Here, we report the structure of the SCL core complex $(SCL:E47)_{bHLH}$:LMO2:$LDB1_{LID}$ bound to DNA. This structure, which details a bHLH:LIM protein association, provides an atomic description of the interactions driving the association of the SCL:E47 heterodimer and its interface with DNA in the presence and in absence of LMO2. Analysis of the structure uncovers unexpected roles for SCL, E47, and LMO2. Functional and biochemical analyses further reveal complex synergies between components of the complex, their cofactors, and DNA targets. Together, these results deepen our understanding of how higher-order, hematopoietic-specific complexes form and how tissue-specific gene expression programs might be regulated. Importantly, the structure, which delivers insights into the oncogenic properties of the complex, will drive the design of small molecules for targeted inhibition of oncogenic processes in T-ALL.

RESULTS

Overview of the Structure

The crystal structures of the $(SCL:E47)_{bHLH}$ heterodimer and that of $(SCL:E47)_{bHLH}$ complexed to LMO2:$LDB1_{LID}$, both bound to a 11 bp blunt-ended E box DNA, were determined at 2.8 and 2.9 Å resolution, respectively (Table 1; Figures 1 and S1C). DNA-bound $(SCL:E47)_{bHLH}$ is topologically similar to previously reported bHLH structures, such as the $(E47:E47)_{bHLH}$ homodimer and $(E47:NeuroD1)_{bHLH}$ heterodimer (Ellenberger et al., 1994; Longo et al., 2008). The basic domains contact the major groove of the DNA (Figure S1A), and the HLH domains, comprising helices 1 and 2 (H1, H2) joined by a loop region, form a parallel left-handed four-helix bundle (Figure 1A). H1 in SCL is four amino acids shorter at its C terminus compared to E47. The large side chain of Y235 projecting from the side of H2 prevents formation of a longer H1 in SCL, whereas in E47, the longer helix is accommodated by residue V595 in the same position (Figure 1A). Because the valine and the tyrosine are conserved residues respectively among class I and class II bHLH proteins (Figure 1E), it is likely that the length of H1 is a characteristic of each bHLH class. The length of H1 impacts the length of the loop: the size and amino acid composition of the loop are poorly conserved across different classes. On the other hand, sequence conservation in this region between SCL, Tal2, Lyl1, Hen1 and Hen2 (all class II proteins binding to

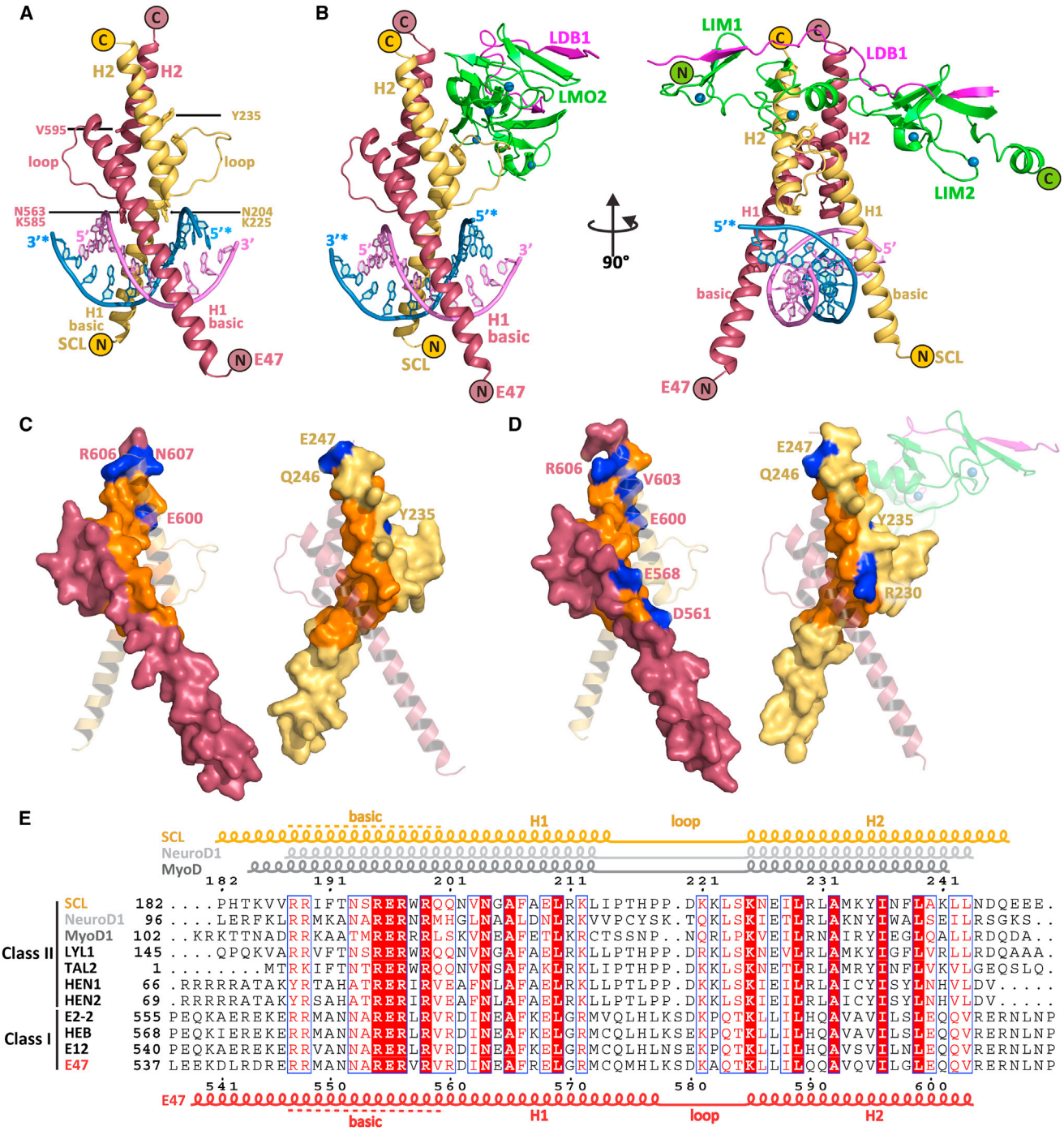

Figure 1. Structures of the DNA-Bound SCL:E47$_{bHLH}$ and SCL:E47$_{bHLH}$:LMO2:LDB1$_{LID}$

(A and B) Ribbon diagrams of the DNA-bound (SCL:E47)$_{bHLH}$ (A) and (SCL:E47)$_{bHLH}$:LMO2:LDB1$_{LID}$ (B) structures. SCL$_{bHLH}$, E47$_{bHLH}$, LMO2, and LDB1$_{LID}$ are colored in yellow, raspberry, green, and magenta, respectively, whereas the DNA strands are shown in blue and pink. basic, basic DNA-binding helix.

(C and D) Surface and ribbon diagram representations showing the residues involved in the HLH interface in the (SCL:E47)$_{bHLH}$ (C) and the (SCL:E47)$_{bHLH}$:LMO2:LDB1$_{LID}$ (D) structures. Hydrophobic residues are colored in orange; residues involved in hydrogen bonding in dark blue.

(E) Sequence and structural alignment of bHLH domains of SCL, E47, and other class I and class II bHLH proteins. Conserved amino acids are drawn in red boxes; similar amino acids are in red type. Secondary structures of the bHLH domain of SCL, NeuroD1, MyoD, and E47 are depicted in yellow, light gray, dark gray, and red, respectively. UniProtKB sequence accession numbers are P17542 (SCL), Q13562 (NEUROD1), P15172 (MYOD1), P12980 (LYL1), Q16559 (TAL2), Q02575 (HEN1), Q02577 (HEN2), P15884 (E2-2), Q99081 (HEB), and P15923 (E12 and E47). PDB accession numbers are 2YPB (SCL and E47), 2QL2 (NeuroD1), and 1MDY (MyoD).

See also Figures S1 and S2 and Table S1.

LMO proteins; Aoyama et al., 2005; Manetopoulos et al., 2003; Wadman et al., 1994; Figure 1E) suggests a role for the loop in directing binding preference to LMO partner proteins.

The three-dimensional structure of DNA-bound (SCL:E47)$_{bHLH}$:LMO2:LDB1$_{LID}$ reveals how LMO2 binds to the bHLH heterodimer (Figure 1B). The interactions occur almost exclusively through the first LIM domain of LMO2 (LIM1), H2 of SCL and E47, and the loop of SCL. Only two residues (R100, R102) of the second LIM domain (LIM2) interface with (SCL:E47)$_{bHLH}$. Our previous work reporting the structure of LMO2:LDB1$_{LID}$ (El Omari et al., 2011) highlighted the functional importance of the LIM interdomain flexibility mediated through hinge residue F88. The structure of (SCL:E47)$_{bHLH}$:LMO2:LDB1$_{LID}$ supports this conclusion because the LIM domains undergo a hinge rotation around F88 to bind to (SCL:E47)$_{bHLH}$ (Figure S2).

The (SCL:E47)$_{bHLH}$ Heterodimer Interface

Interactions between bHLH proteins are driven by the strength of their homodimeric versus heterodimeric association. The composition of the buried hydrophobic interface of (SCL:E47)$_{bHLH}$ (orange in Figure 1C; Table S1) is very similar to that of (E47:E47)$_{bHLH}$ (Ellenberger et al., 1994). Sequence conservation of these hydrophobic residues across class I and II bHLH proteins (Figure 1E) suggests structural conservation of the HLH interface. Hydrophilic amino acids (Table S1, light-blue residues) are also found at the bHLH interface but are less conserved and might reflect heterodimerization specificity between different bHLH proteins. One salt bridge and two hydrogen bonds are formed between H2 residues from SCL and E47 in the (SCL:E47)$_{bHLH}$ structure (dark-blue residues in Figure 1C and Table S1): E247:R606, Q246:N607, and Y235:E600, respectively. Because this latter bond is also seen in (NeuroD1:E47)$_{bHLH}$ (Longo et al., 2008), it might be a common feature of all E47 heterodimers because Y235 is conserved among class II bHLH factors (Figure 1E). The E47 homodimer interface contains two symmetrically positioned hydrogen bonds between E600 of H2 from one monomer and H576 in H1 from the other monomer (Ellenberger et al., 1994). No structure is available for the SCL homodimer, but it is very likely that, like in the MyoD homodimer structure (Ma et al., 1994), no hydrogen bonds are formed between monomers, a characteristic that may decrease the stability of class II homodimers. In our hands, SCL$_{bHLH}$ homodimers could not be purified in isolation, whereas stable (SCL:E47)$_{bHLH}$ heterodimers could be formed by coexpressing the two proteins; moreover, SCL homodimers are not described in vivo. These results are consistent with the view that E47:SCL is thermodynamically favored in solution because SCL homodimers are disfavored, as previously suggested for NeuroD1 (Longo et al., 2008). Moreover, association constants have been measured for (E12:E12)$_{bHLH}$, (SCL:SCL)$_{bHLH}$, and (SCL:E12)$_{bHLH}$ with a respective K_Δ of 7.4×10^4, 5.3×10^6, and 3.5×10^7 M^{-1} (Ryan et al., 2008), in agreement with the fact that the heterodimer is more stable than individual homodimers. Note that E2A isoforms E47 and E12 are highly homologous in their bHLH region and show 100% conservation in the residues at the interface with SCL and DNA, suggesting very similar structural and biophysical properties (see Figure 1E and Table S1).

Analysis of SCL:E47 structures in the presence of LMO2 reveals the structural impact of LMO2 on heterodimerization. Binding of LMO2 only slightly increases the overall surface buried in the (SCL:E47)$_{bHLH}$ interface (2,185 versus 2,149 Å^2) (Figure 1D). However, a substantial gain is seen in the network of hydrogen bonds. Two new hydrogen bonds (Q246:V603 and R230:D561) and one salt bridge (R230:E568) are now formed (dark blue, Figure 1D; Table S1). Interestingly, two of these new bonds connect H2 of SCL to H1 of E47. Salt bridges have been implicated in the dimerization strengths and preferences of bHLH proteins (Shirakata et al., 1993). By inducing the formation of new hydrogen bonds and salt bridges, LMO2 binding is likely to increase the heterodimerization affinity over E47 homodimerization. Moreover, the reported in vitro association constant of LMO2 to the heterodimer ($\sim$1.8 10^8 M^{-1}) (Ryan et al., 2008) is much stronger than either homo- or heterodimeric association constants (see above), suggesting an important function for LMO2 in ''locking'' the heterodimer structure.

Interaction of SCL:E47$_{bHLH}$ with DNA

(SCL:E47)$_{bHLH}$ is oriented on the E box motif (forward, 5′-C^{1}A^{2}G^{3}A^{4}T^{5}G^6-3′; reverse, 5′-C^1*A^2*T^3*C^4*T^5*G^6*-3′); SCL binds the CAG half-site and E47 the C*A*T* half-site (Figures 2A, 2B, and S1A). Compared to other bHLH structures, (SCL:E47)$_{bHLH}$ makes fewer overall contacts to DNA. The conserved glutamates E555 (E47) and E196 (SCL) form the only specific interactions with the DNA bases. Notably, E555 of E47 contacts the DNA loosely, making one putative hydrogen bond to N4 of C^{1*} (3.6 Å) and sitting at a distance of 3.9 Å from T^5. The orientations of the conserved glutamate side chains are stabilized by a salt bridge with conserved neighboring arginines (SCL R199, E47 R558). All other interactions with DNA are through the phosphate backbone. Four bonded contacts are formed by SCL and three by E47 (Figure 2C). The side chains of conserved lysines at the tip of H2 (SCL K225, E47 K585; see Figure 1A), capped by the phosphate of C$^{1/1*}$, are stabilized by a salt bridge to conserved asparagines on H1 (SCL N204, E47 N563; see Figure 1A), suggesting a bHLH canonical mechanism to anchor the position of H1 and H2 to each other and to the DNA. All the other contacts with the DNA backbone are nonbonded contacts.

In our structure, the (SCL:E47)$_{bHLH}$ heterodimer does not directly contact the central bases of the E box motif, although water-mediated contacts cannot be excluded at the resolution of our structures. This is in contrast to the specific contacts observed in the (E47:E47)$_{bHLH}$:DNA structure (Ellenberger et al., 1994). Indeed, the two E47 monomers, crystallized in the presence of the consensus sequence for E47 (CACCTG), contacted the DNA asymmetrically: the ''nonspecific'' monomer interacted with the C*A*G* half-site through binding of E555 to the base of C^{1*}; the ''specific monomer'' interacted with CAC through binding of R556 to the central base G^{4*}. In the (SCL:E47)$_{bHLH}$ structure, both the E47 and the SCL basic helices adopt the nonspecific conformation. Although a specific conformation is not observed in the crystal structure itself, this does not preclude the possibility of a rotation in solution of the E47 R556 side chain to contact the central base A^4. The basic helix of SCL appears however to be less flexible and less likely to adopt the specific conformation (i.e., rotation of the R197

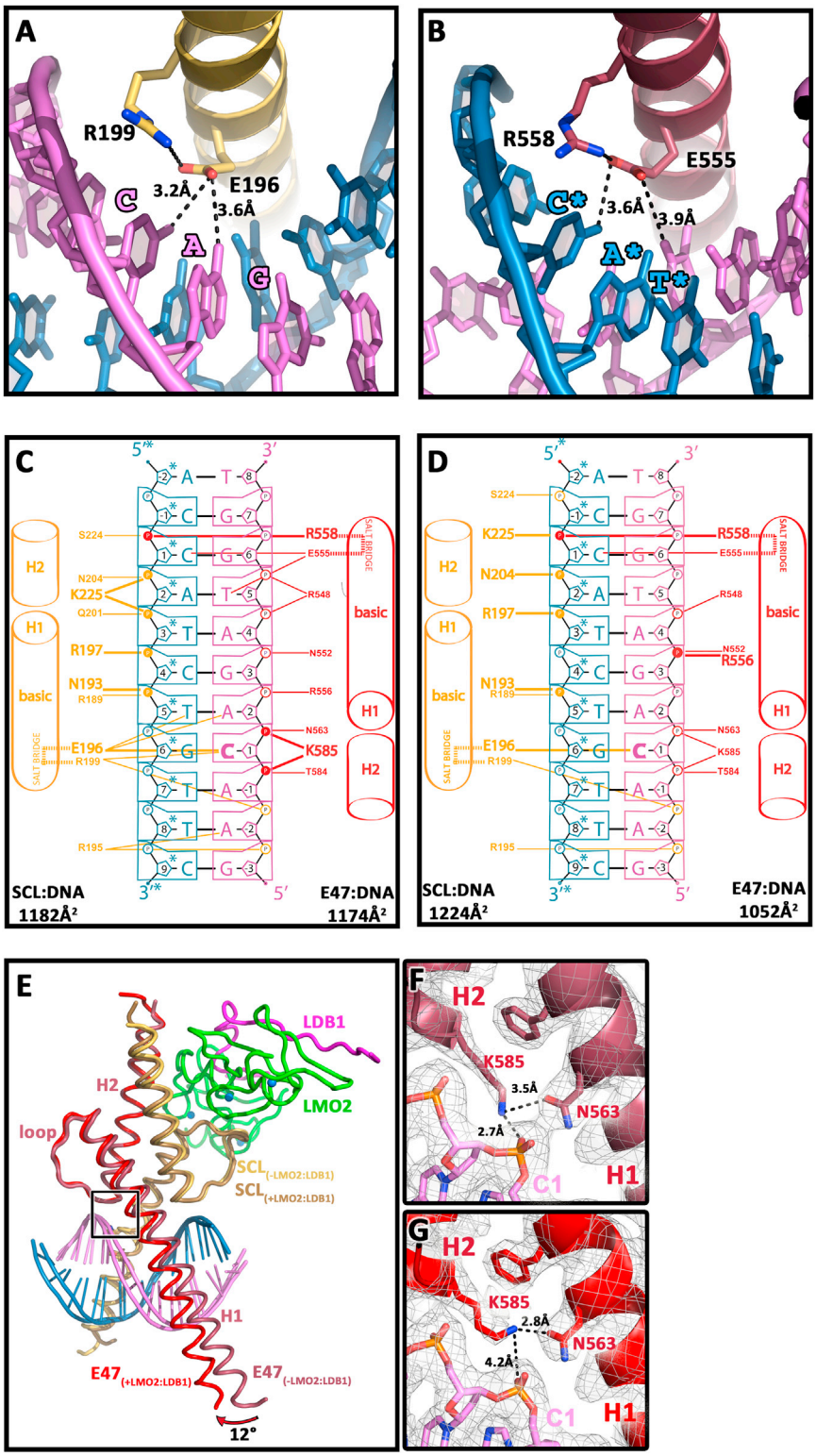

Figure 2. The DNA Recognition Interfaces

(A and B) Detailed view of SCL_{bHLH} (A) and $E47_{bHLH}$ (B) interactions in the major groove.

(C and D) Schematic diagram of detailed protein-DNA contacts of the $(SCL:E47)_{bHLH}$ structure (C) compared with the contacts of the $(SCL:E47)_{bHLH}$:$LMO2:LDB1_{LID}$ complex (D) using NUCPLOT (Luscombe et al., 1997). Hydrogen bonds (within 3.5 Å) are represented as thick lines to the backbone (filled circles) or to the bases (bold letter). van der Waals contacts (within 3.9 Å) are represented as thin lines.

(E) Subdomain motions of the bHLH domains induced by LMO2 binding. Ribbon diagrams of the DNA-bound $(SCL:E47)_{bHLH}$ and $(SCL:E47)_{bHLH}$:$LMO2:LDB1_{LID}$ structures superimposed based on H2 of E47. The angle of rotation required to superimpose the basic and H1 helices of E47 is 12°. Color coding is yellow and gold (SCL), raspberry and red (E47) for the $(SCL:E47)_{bHLH}$ and $(SCL:E47)_{bHLH}$:$LMO2:LDB1_{LID}$ structures, respectively. The black rectangle corresponds to the region shown in (F) and (G).

(F and G) Close-up views of the area around E47 residues K585 and N563, showing 2Fo – Fc electron density maps contoured at 1 s level (gray) and distances between residues and DNA. (F) $(SCL:E47)_{bHLH}$ structure. (G) $(SCL:E47)_{bHLH}$:$LMO2:LDB1_{LID}$ structure. Proteins and DNA are colored as in (E).

See also Figure S1A.

side chain to contact the base of C^{4*}). Indeed, Q201 in SCL contacts the phosphate backbone at T^{3*}, whereas the equivalent residue in E47 (R560) does not, giving more flexibility to the protein-backbone interaction (Figure 2C). Similarly, the conserved glutamate E196 in SCL contacts the bases of C^{1}, A^{2}, and T^{5*} more closely than the equivalent residue E555 in E47 (Figures 2A–2C), seemingly locking SCL in its unspecific conformation. It seems therefore likely that if $(SCL:E47)_{bHLH}$ is able to bind E boxes specifically, its orientation on DNA is dictated by E47.

Binding of LMO2 to SCL:E47 imposes unsuspected structural constraints affecting heterodimer:DNA interactions. As discussed above, binding of LMO2 tightens the heterodimerization strength by increasing the network of hydrogen bonds, including the creation of a salt bridge between H2 of SCL (R230) and H1 of E47 (E568). The creation of these new bonds following LMO2 binding is likely to be caused by subtle rearrangements occurring to H2 of SCL, the main interface of interaction of the heterodimer with LMO2 (see below). The creation of the new bonds and the subtle conformational changes at the heterodimer interface induce movements of the H1 helices and, consequently, of the basic domains. Specifically, the hinge angle between H1 and H2 of E47

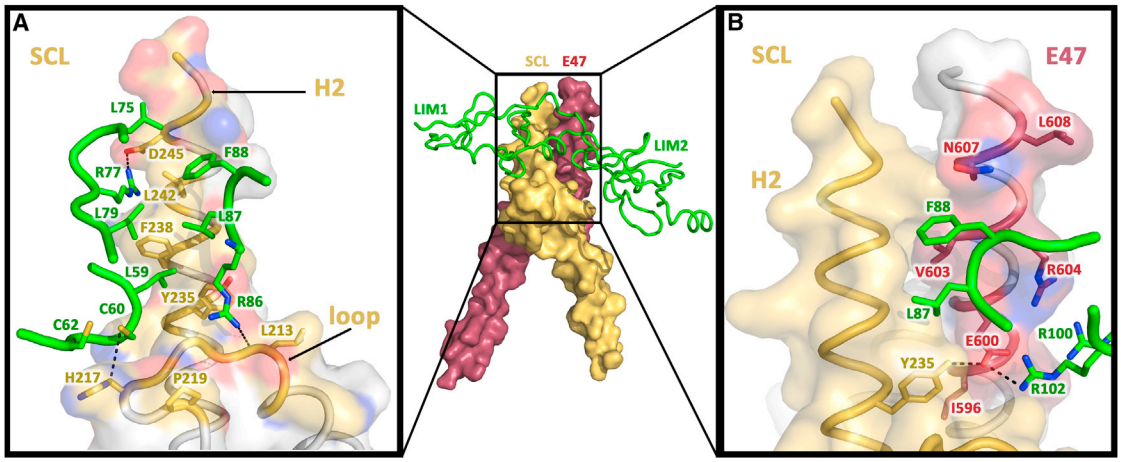

Figure 3. The SCL:E47$_{bHLH}$:LMO2 Interface

(A and B) Surface and ribbon representation of the interface between LMO2 and, respectively, SCL$_{bHLH}$ (A) and E47$_{bHLH}$ (B). Residues found at the interface are depicted in a stick representation. Interaction surfaces of SCL and E47 are colored in yellow and red, respectively, with overlaid electrostatic potential (blue indicates positive; red is negative) as calculated by the program PyMOL (DeLano, 2002). See also Figure S1C.

undergoes a large rotation of about 12° (Figure 2E). This causes a shift in the bond between the conserved lysine K585 at the tip of E47 H2 and conserved asparagine N563 in E47 H1 (Figures 2F and 2G), a loosening of the contacts between K585 and backbone DNA (from 2.7 to 4.2 Å, Figures 2D, 2F, and 2G), and a decrease in the overall surface buried between E47 and DNA (~120 Å^2). The overall network of bonded and nonbonded contacts between the LMO2-bound heterodimer and DNA decreases in size compared to the nonbound form (from 73 to 56 contacts). As a result, binding of LMO2 is predicted to destabilize the (SCL:E47)$_{bHLH}$:DNA interface. Of note, in vitro measurement of the DNA-binding affinities shows that the (SCL:E12)$_{bHLH}$:LMO2:LDB1$_{LID}$ complex binds palindromic and nonpalindromic E box motifs with lower affinity (between 3.7- and 1.5-fold) than (SCL:E12)$_{bHLH}$ (K_Δ = 4 × 10^5 M^{-1} versus K_Δ = 1.5 × 10^6 M^{-1} for CAGCTG, and K_Δ = 1.7 × 10^6 M^{-1} versus K_Δ = 2.5 × 10^6 M^{-1} for CAGGTG) (Ryan et al., 2008).

Interaction of SCL:E47$_{bHLH}$ with LMO2:LDB1$_{LID}$

Analysis of the (SCL:E47)$_{bHLH}$:LMO2:LDB1$_{LID}$ structure defines the SCL:LMO2 interface (Figure 3A). Despite a relatively high reported association constant (Ryan et al., 2008), the SCL:LMO2 complex buries a surface area of only 1,232 Å^2, which is at the lower end of the "standard size" area (1,200–2,000 Å^2) buried in heterodimeric protein-protein complex (Lo Conte et al., 1999). The SCL:LMO2 interaction occurs through residues localized within a small region of their surfaces, mainly through H2 and loop of SCL and LIM1 of LMO2 (Figure 3A), involving an approximately equal number of hydrophobic and hydrophilic interactions. The surface-exposed hydrophobic residues of SCL (F238, L242, L213) and LMO2 (L59, L79, L87, F88) are the main stabilizers of the interface, whereas two hydrogen bonds (L213:R86, H217:C60) and a salt bridge (D245:R77) provide specificity.

Importantly, LMO2 interacts with E47. However, the surface area buried in this interface is very small (622 Å^2). Despite the limited number of hydrophobic and hydrophilic interactions centered around residues F88 (LMO2 LIM1) and R604 (E47), a robust salt bridge is formed between R102 (LMO2 LIM2) and E600 (E47), which in turn forms a hydrogen bond with Y235 (SCL) (Figures 3B and S1B). A second arginine in LMO2, R100, is also in close proximity of E600 (4.5 Å), creating a highly charged binding groove. Potentially, this interaction is the driving force for the change in the SCL:E47 interface packing.

Functional Identification of the Residues Involved in the SCL:LMO2 Interaction

To probe the relationship between the observed crystal structure interactions and their biological role, we generated a series of mutations predicted to perturb the interfaces between SCL, LMO2 and E47.

We first tested the capacity of LMO2 interface mutations (LMO2-L59G, R77A, R86A, F88D, and R100A/R102A) to bind to wild-type (WT) SCL in two-hybrid luciferase reporter assays (Figures 4A and S3A). The L59G mutation increased the luciferase activity, suggesting that other exposed hydrophobic residues compensate for the absence of this residue. With R77A and R86A, luciferase activity was reduced by 60% and 80%, respectively, pointing at a specific function for the residues mediating the hydrogen-bonding network. No activation was observed with the F88D hinge mutant, confirming our previous conclusions that the SCL:LMO2 interaction is strongly affected when the rotation movement between the two LIM domains of LMO2 is restricted (El Omari et al., 2011). Finally, the LIM2 domain mutant R100A/R102A was tested for its interaction with SCL. R102 is the only residue of LMO2 in hydrogen bond contact with E47; moreover, both R100 and R102 are not conserved in other LMOs, suggesting an LMO2-specific function. Surprisingly, the mutation increased the luciferase activity, suggesting that the function of these residues is not to provide specificity for the interaction but, perhaps, to lock the relative positions of LIM1 and LIM2 when binding to SCL:E47.

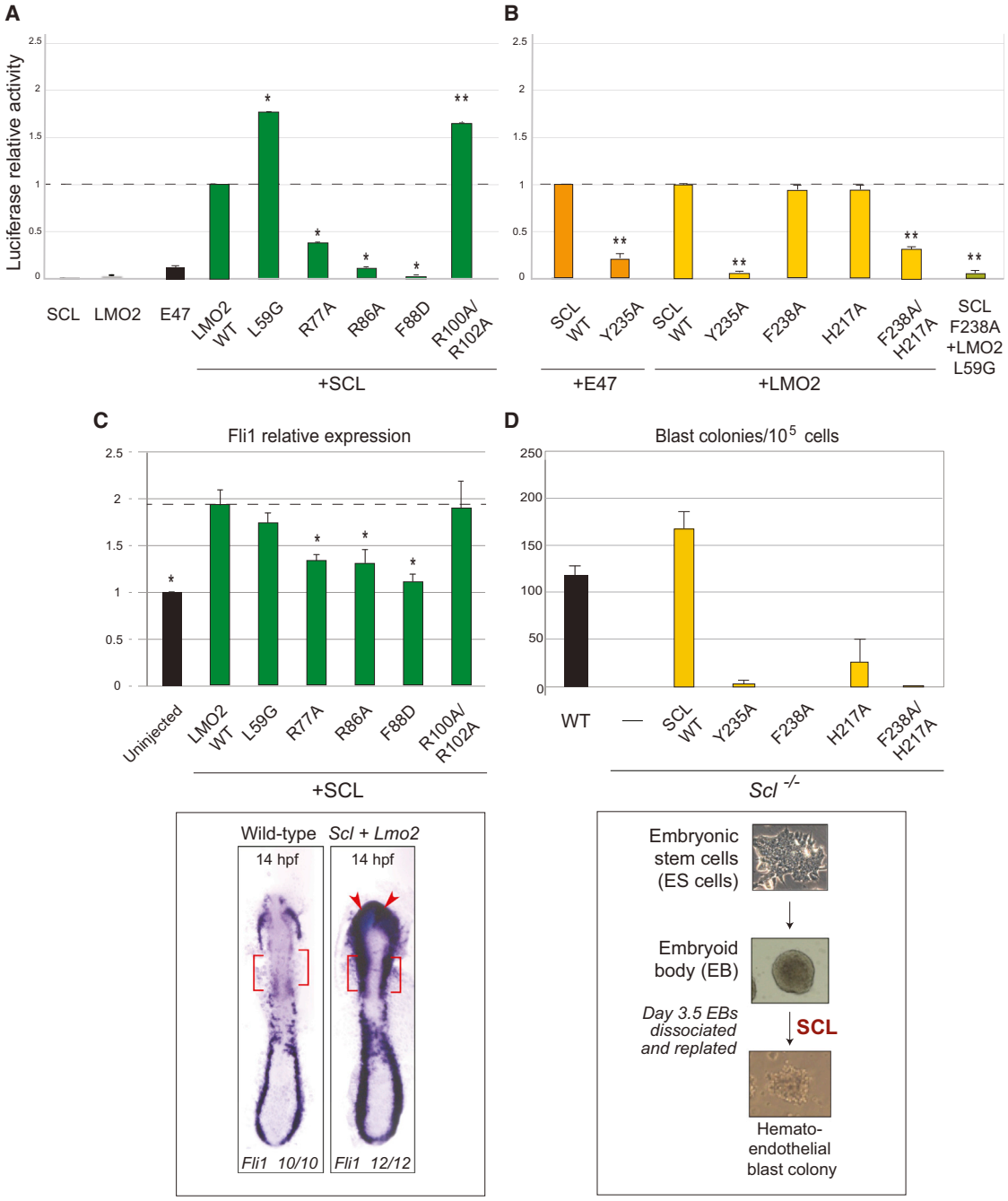

Figure 4. Identification of Functionally Critical Residues at the SCL:E47:LMO2 Interface

(A and B) Luciferase-based mammalian two-hybrid assays were conducted in HEK293 cells with VP16-SCL, Gal4-LMO2, and Gal4-E47 fusion proteins. (A) WT and mutant LMO2 proteins are coexpressed with WT SCL. (B) WT and mutant SCL proteins are coexpressed with WT E47, WT LMO2, or LMO2-L59G. Luciferase activities resulting from interactions between WT proteins are set to 1. Error bars represent the SE from three biologically independent.experiments.

(C) Gain-of-function studies in zebrafish embryos. One-cell-stage embryos were coinjected with WT *Scl* mRNA and WT or mutant *Lmo2* mRNA. Expansion of the field of expression of the hemangioblast marker *Fli1* was monitored by reverse-transcription qPCR (top) or whole-mount in situ hybridization (bottom) on ten-somite-stage embryos (14 hpf). Error bars represent the SE from four biologically independent.experiments. Brackets indicate cardiac mesoderm; arrows point to head mesoderm. Anterior to the top.

(D) Phenotypic rescue of hematopoiesis from $Scl^{-/-}$ ES cells is detected using the BL-CFC assay (bottom). Blast colonies were generated upon in vitro differentiation of WT ES cells and $Scl^{-/-}$ ES cells untransfected (–) or rescued with WT or mutated versions of SCL, as indicated (top). Error bars represent the SE from three biologically independent.experiments.

*$p < 0.05$; **$p < 0.01$.

See also Figure S3.

To analyze the functional consequences of these mutations in vivo, we performed gain-of-function studies in zebrafish embryos (Gering et al., 2003). Coinjection of WT *HA-Scl* and *Gfp-Lmo2* mRNA leads to an expansion of *Fli1*-expressing hemangioblasts (mesodermal precursors of endothelial and hematopoietic cells) in the head, cardiac territory, and posterior domain of the embryos (Figure 4C, bottom panel); quantitative PCR (qPCR) at the ten-somite stage shows a 2-fold increase in *Fli1* expression levels when compared to WT embryos (Figure 4C, top panel). When expressed with WT SCL, LMO2 L59G and R100A/R102A did not affect *Fli1* expression, whereas the mutants R77A, R86A, and F88D led to a significant reduction in *Fli1* expression when compared to WT LMO2 (Figures 4C and S3B), in agreement with the two-hybrid assay (Figure 4A).

We then assessed the role of three residues in SCL for interaction with E47 and LMO2 and for their function in hematopoietic development. Y235, in H2, lies at the interface between SCL and E47. H217 and F238, at the interface between SCL and LMO2, are conserved in hematopoietic bHLH proteins (LYL1, TAL2), but not in neuronal and muscle proteins (NeuroD1, HEN1, HEN2, and MyoD, Figure 1E), suggesting a hematopoietic-specific function. When tested in the two-hybrid assay, interaction of SCL Y235A with E47 and LMO2 was greatly impaired (Figures 4B and S3A), confirming (1) the importance of Y235 for heterodimerization, and (2) the necessity of heterodimer formation for SCL to interact with LMO2. Individual mutations H217A and F238A did not affect the SCL:LMO2 interaction; however, when combined (H217A/F238A), interaction was reduced by two-thirds. Combining SCL F238A and LMO2 L59G mutations totally abolished the interaction.

To test these mutations functionally, we performed phenotypic rescue assays of $Scl^{-/-}$ ES cells. WT ES cells differentiate into cellular structures called embryoid bodies (EBs) that contain hematopoietic precursors giving rise to blood lineages. Using this assay, we have previously shown that the SCL:LMO2 interaction was required for the development of primitive erythropoiesis (Schlaeger et al., 2004). In contrast, here, we assessed the formation and development of blast colony-forming cells (BL-CFCs or hemangioblasts), the earliest clonal ES cell-derived mesodermal precursors of blood and endothelial cells (Choi et al., 1998) (Figure 4D). This process requires SCL activity (D'Souza et al., 2005); therefore, blast colonies were not observed from $Scl^{-/-}$ cells. Blast colony-forming potential was fully rescued upon expression of a WT *Scl* cDNA into $Scl^{-/-}$ ES cells (Figures 4D and S3C). The single (Y235A, F238A, H217A) and double (F238A/H217A) mutants all abolished SCL activity, highlighting the impact of key, single mutations on the function of the multiprotein complex in vivo.

Altogether, these data define residues critical for formation of a stable SCL:E47:LMO2 interface and demonstrate the requirement for SCL heterodimerization and interaction with LMO2 for its earliest role in hematopoietic development.

Building Higher-Order Multiprotein Complexes

In erythroid cells, the SCL core complex interacts with GATA1 and cofactors that confer transcriptional activity. In order to understand how these higher-order complexes assemble, we evaluated the contribution of the components of the complex to the recruitment of known interacting coactivators (histone acetyltransferases PCAF and P300), corepressors (ETO2 and histone deacetylase mSIN3A), and histone demethylase LDS1. Guided by our structural and functional analyses, we selected point mutations that precisely disrupt heterodimerization (SCL Y235A) and interaction of SCL with LMO2 (SCL H217A/F238A and LMO2 R86A); we also tested mutation of the hinge residue in LMO2 (F88D). We expressed the corresponding mutant forms of SCL or LMO2 as biotinylated proteins (bio-SCL or -LMO2) in erythroid MEL cells (Schuh et al., 2005). Upon streptavidin-affinity pull-down, copurified proteins were identified.

Members of the core complex, GATA1, and all the cofactors tested were efficiently pulled down with WT bio-SCL (Figure 5A, left panels, PD fraction). SCL Y235A prevented interaction with E47 and all core members and with mSIN3A, PCAF, ETO2, and P300, confirming that heterodimerization is essential for SCL activity. Interestingly, interaction between SCL Y235A and LSD1 was preserved, suggesting a different mechanism of recruitment. The SCL H217A/F238A mutations abolished interaction with LMO2 and, consequently, with LDB1 and GATA1, but not with the cofactors; their recruitment therefore does not require interaction with LMO2 and LDB1 and is likely to occur via the heterodimer.

WT bio-LMO2 (Figure 5A, right panels, PD fraction) efficiently interacted with the core complex, GATA1, and the cofactors. The R86A mutant, unable to interact with SCL and E47, retained binding to LDB1 and GATA1, but not to the corepressors mSin3A, ETO2, and LSD1. It could however still interact with the coactivators PCAF and P300. When the conformational flexibility of LMO2 was impaired (F88D), interaction with SCL and E47 was abolished but not with LDB1 nor GATA1, indicating that these latter interactions can occur independently of the hinge movement and, in contrast to recent suggestions by Dastmalchi et al. (2012), confirming that mutation of F88 does not overall affect the structure of LMO2. mSIN3A, PCAF, ETO2, and P300 were not able to interact with LMO2 F88D; however, we detected an interaction with LSD1. Altogether, this suggests that interaction of LMO2 with mSin3A and ETO2 relies on binding to the heterodimer, whereas interaction with the coactivators can occur independently, yet it requires the hinge movement. The LMO2:LSD1 interaction does not require SCL:E47 but only occurs when LMO2 conformation is locked, possibly through GATA1 and its partner FOG1 (Snow and Orkin, 2009).

DISCUSSION

This report details at the atomic and functional level the exquisite interplay of interactions between the bHLH proteins SCL and E47 and the LIM protein LMO2, thereby delivering mechanistic insights into hematopoietic specification and leukemogenic transformation. Given the importance of bHLH:LIM associations in development and differentiation processes, this provides general principles that will apply to other tissue-specific complexes.

LMO2 Stabilizes the SCL:E47 Association but Weakens the Interaction of the Heterodimer with DNA

In addition to their association into heterodimers, SCL and E2A bHLH domains can exist in vitro as homodimers (although

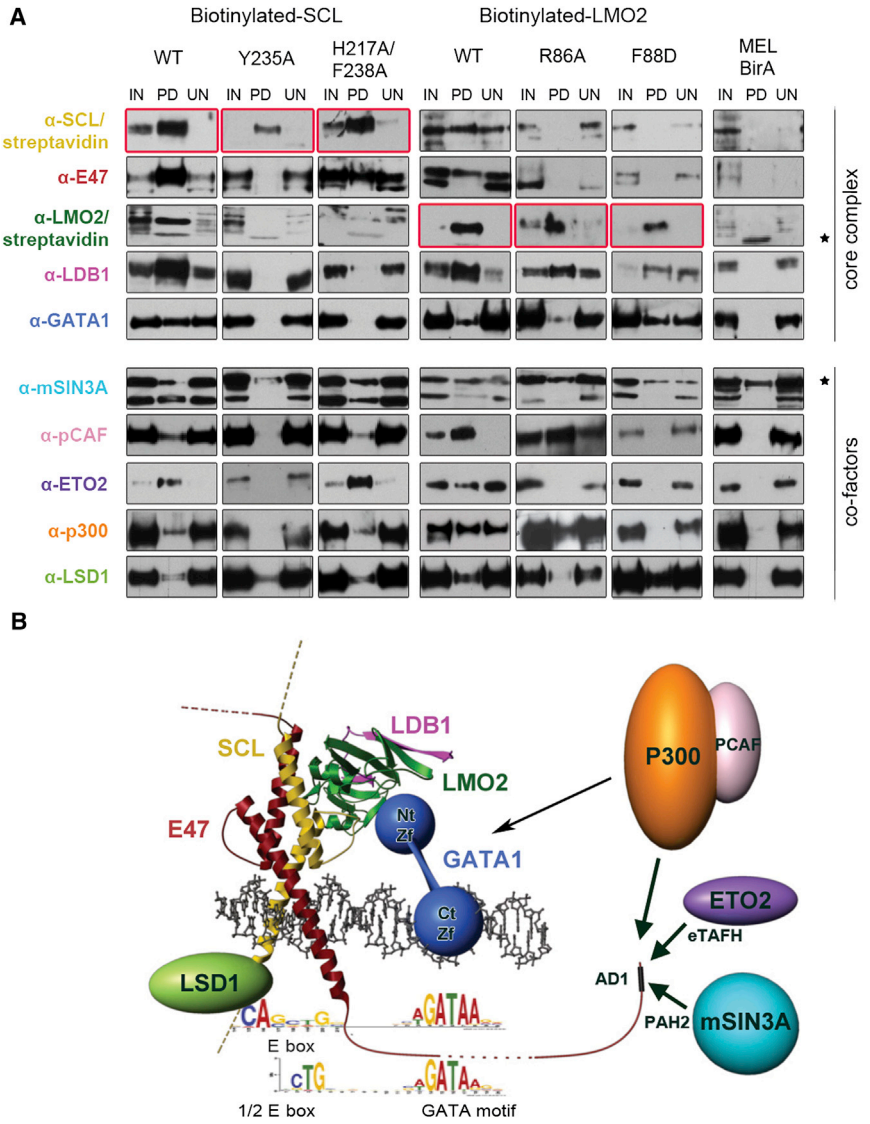

Figure 5. Building a Network of Interactions between SCL:E47:LMO2:LDB1 and GATA1, Corepressors, Coactivators, and Chromatin-Remodeling Proteins

(A) Multiprotein complexes containing bio-SCL (WT, Y235A, or H217A/F238A) or bio-LMO2 (WT, F88D, or R86A) were affinity purified from MEL cell nuclear extracts and analyzed by western blotting for the presence of other members of the core complex and known cofactors. Antibodies used are indicated. Streptavidin-HRP was used to detect bio-SCL or LMO2 (red-framed western blots). MEL BirA, negative control. The asterisk (*) indicates unspecific bands.

(B) Model representing the pentameric complex bound to its recognition motifs (E box-GATA or 1/2 E box-GATA; Kassouf et al., 2010) and its interactions with known cofactors as suggested by our biochemical study. The structure of the quaternary complex bound to DNA is depicted as in Figure 1B. The Nt portion of E47 and the activation domain AD1 are schematically represented as a red ribbon to be able to position ETO2, P300, PCAF, and mSin3A. See Discussion for details. See also Figure S4.

to date, only E2A homodimers have known biological functions). Our results provide a structural explanation for the hetero/homodimerization preferences and for the modulation of these affinities upon LMO2 binding. As detailed in the Results section, we suggest that the SCL:E47 heterodimerization interface is more stable than either individual homodimerization interfaces and that binding of LMO2 induces subtle conformational changes resulting in the creation of new hydrogen bonds.

Furthermore, conserved hydrogen bonds, found in *cis* between H1 and H2 of both SCL and E47 and in *trans* between SCL H2 and E47 H1, suggest a clear link between the angle of the HLH dimerization domain and the DNA contacts available to the basic domains. Binding of LMO2 alters the dimerization angle and weakens the interactions between the basic helices of SCL:E47 and DNA.

Therefore, our structural data reveal how LMO2 strengthens the SCL:E47 interaction while weakening its association with DNA. As detailed below, this provides a powerful level of regulation of complex formation and transcriptional activities.

Implications for Leukemogenesis and Normal Hematopoiesis: The Sequestration Model

Ectopic expression of oncogenic TFs in T cell precursors causes T-ALL by inhibition of E2A:E2A target gene expression and aberrant activation of novel targets (Palii et al., 2011; Sanda et al., 2012). Our work provides strong evidence for the E2A-sequestration model in SCL-expressing T-ALL through formation of SCL:E2A heterodimers. Simultaneous expression of LMO2 reinforces this effect and weakens the DNA:heterodimer interactions. Consequently, LMO2 binding is likely to alter expression of E2A:E2A targets by facilitating the displacement of SCL:E2A from DNA. The complex can then be relocated to alternative genomic sites by association with other DNA-binding factors such as RUNX1, ETS1, and GATA3, as recently suggested by the reported association of SCL-binding events with Runx, Ets, and Gata motifs in T-ALL cells (Palii et al., 2011; Sanda et al., 2012) (Figure 6). This is corroborated by transcriptome analyses showing expression of different gene sets in T-ALL cells upon knockdown of SCL (Palii et al., 2011). Therefore, our structure explains at the atomic level how ectopic expression of SCL and LMO2 overall alters the gene repertoire in T cells to ultimately lead to oncogenic transformation.

The sequestration model might also help explain the roles of SCL and LMO2 in hematopoietic specification. Coexpression of SCL and LMO2 in mesodermal precursors would favor

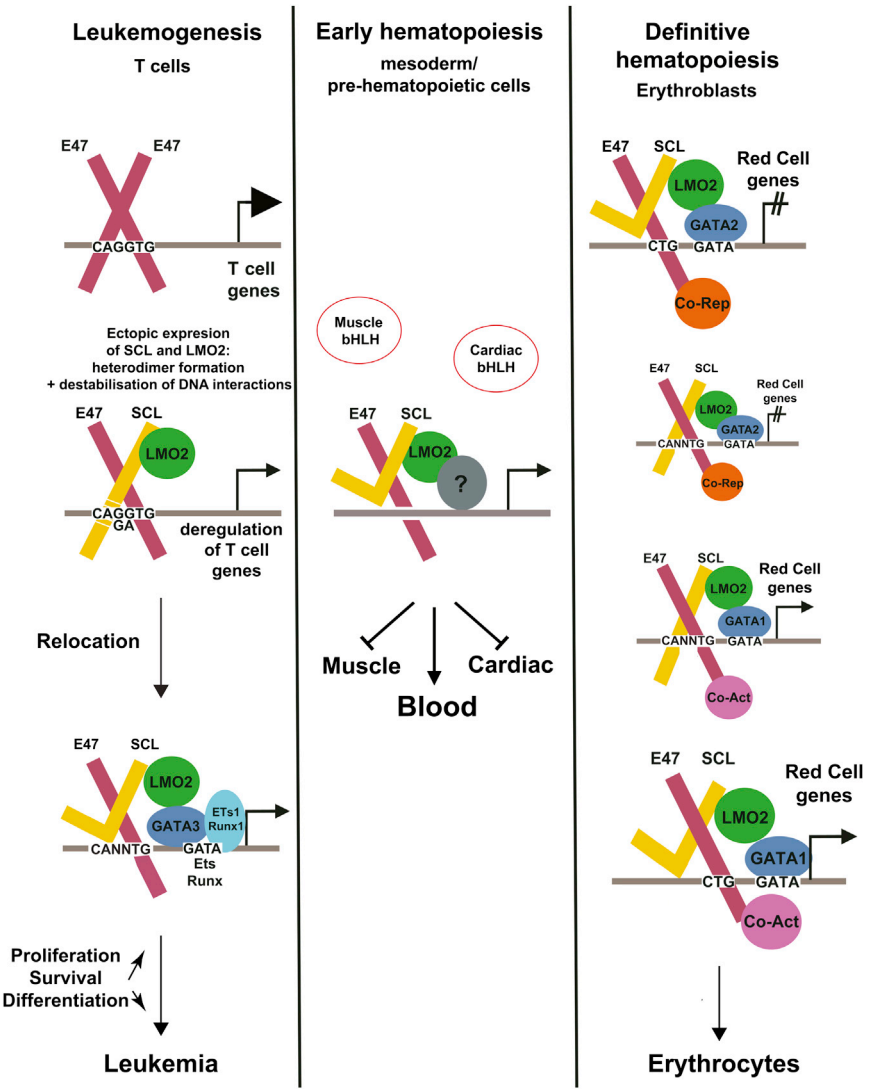

Figure 6. Model of SCL-Mediated Network of Interactions in Leukemogenesis and Early and Definitive Hematopoiesis

The structural and functional data integrated into this model are discussed in the text. The consensus E box motifs are from Kassouf et al. (2010), Palii et al. (2011), and Soler et al. (2010). Although individual motifs have been reported, composite motifs have not been described in T cells. The DNA motifs and additional TFs recruiting the complex in mesoderm/pre-hematopoietic cells have not been identified. The "kinked" representation of SCL symbolizes its reported DNA-binding independent activities. The different sizes of the complexes in erythroid cells reflect the proportion of SCL-containing binding events on composite motifs associated with repressed and activated target genes; for details, see Kassouf et al. (2010).

heterodimerization of E2A with SCL as opposed to other bHLH regulators, thereby imposing a hematopoietic fate at the expense of alternative cell fates. This agrees with recent studies showing ectopic cardiac development in the absence of SCL (Schoenebeck et al., 2007; Simões et al., 2011; Van Handel et al., 2012) (Figure 6).

In line with this, our phenotypic rescue of ES cell-derived BL-CFC complemented by gain-of-function studies in zebrafish provides a demonstration that specification of mesodermal precursors toward the hematopoietic lineage critically depends on the activity of a functional SCL:E47:LMO2 complex. Further characterization will reveal some of the mechanisms underlying the earliest stages of blood development.

Structural Basis for the DNA-Binding-Independent Functions of SCL

Dissection of the $(SCL\text{-}E47)_{bHLH}$:DNA interface shows weak interactions with nucleic acids. This might have physiological relevance, allowing fine-tuning of the activities of SCL:E47-containing complexes in vivo through interactions with additional DNA-bound TFs. This is also relevant for the DNA-binding independent activities of SCL in hematopoietic specification, in leukemogenesis, and in interacting with some of its erythroid genomic targets. In red cells, LMO2 mediates interactions of SCL:E47 with GATA1, the complex binding a composite CANNTG-(N)9-GATA. A manually docked pentameric complex bound to E box-GATA DNA (Figure S4) shows that the complex can be assembled only if the 30 amino acid linker between the N- and C-terminal domains of GATA1 assumes a semiextended conformation. In this model, and as dictated by the interaction of the heterodimer with LMO2, SCL binds to the GATA distal half-E box and E47 to the proximal one. Remarkably, the occurrence of a novel composite motif missing the distal SCL-bound half-E box (CTG-(N)9-GATA) (Figure 5B) is increased by up to 3-fold in cells expressing a DNA-binding mutant form of SCL (Kassouf et al., 2010).

Therefore, a model emerges both in leukemia and hematopoiesis by which DNA-binding specificity is not driven by SCL:E47 but by its DNA-bound protein partners, the identity of which depends upon the cellular context (Figure 6). This model could explain the lack of interaction of SCL:E47 with the central bases of the E box motif, in contrast to what is seen in the $(E47{:}E47)_{bHLH}$ homodimer. Indeed, whereas we do not rule out the possibility that binding of the specific monomer to central bases could still be achieved in solution, the lack of interaction observed in our crystals could alternatively be a hallmark of bHLH heterodimers. Because recruitment of the SCL complex to DNA essentially relies on interaction with LMO2 and other DNA-binding proteins, a degenerated E box sequence

(CANNTG) would be sufficient to anchor the heterodimer to DNA. In contrast, E47 homodimers do not bind to LMO2 and require additional contacts with DNA for specific genomic target recognition. Altogether, these observations refine our understanding of how specific binding in promoters/enhancers is achieved and reveals an additional level of the mechanistic diversity of TFs.

SCL Directs E47 and Main Cofactors to Hematopoietic-Specific Targets

Our mutational and biochemical analysis uncovers a major role for SCL in directing E47 and associated regulators to DNA (Figure 5B). The regulators ETO2, P300, PCAF, and mSin3A are recruited to the complex by the SCL:E47 heterodimer defective for its interaction with LMO2:LDB1, but not by the heterodimerization-defective form of SCL. This agrees with a previous report mapping the E protein:ETO-related proteins and E protein:P300 interactions to the AD1 domain of E proteins (Zhang et al., 2004). Similarly, PCAF, a P300-associated factor, likely interacts with the AD1 domain. Finally, given the structural similarity of the AD1-binding domain of ETO-related proteins (eTAFH) and the paired amphipathic helix domain (PAH2) of mSIN3A (Plevin et al., 2006), mSin3A also plausibly binds the AD1 domain. Therefore, through both tissue-specific expression and interaction with LMO2, SCL's main function appears to direct E47 and associated cofactors/chromatin remodelers to lineage-restricted target genes.

In contrast, SCL interacts with LSD1 independently of its dimerization to E47, in agreement with their described direct association (Li et al., 2012). In this context, SCL could dimerize with partners other than E proteins, such as HLH ID proteins (Condorelli et al., 1995). SCL:ID complexes might regulate transcription through formation of transient, non-DNA-binding structures sequestering cofactors, such as LSD1, before releasing SCL and interacting proteins for association with E47 into DNA-bound complexes.

LMO2: A Versatile Adaptor Protein

In erythroid cells, LMO2 associates with corepressors mSIN3A and ETO2 only when in a complex with SCL:E47; therefore, not through GATA1. This corroborates previous studies showing that SCL repressor functions are GATA1 independent (and likely to be mediated, at least in part, through interaction with GATA2; Fujiwara et al., 2009), whereas the SCL:GATA1 complex is an activator complex (Tripic et al., 2009) (Figure 6). Moreover, GATA1 repressive complexes do not contain LMO2 (Tripic et al., 2009) and are therefore not associated with SCL:E47. Interestingly, the coactivators P300 and PCAF, although directly bound to E47, can interact with LMO2 independently of its association with SCL:E47. Because most nuclear LMO2 is bound to SCL (this work; Lécuyer et al., 2007), this interaction is likely to be mediated by another TF associated with the complex, such as GATA1. We suggest that acetylation of GATA1 by P300 and PCAF, described as facilitating associations with protein partners and chromatin templates (Lamonica et al., 2006), plays a critical role in the activating functions of the pentameric complex.

The hinge movement is differentially required for the LMO2 interactions with SCL:E47 and GATA1 or for its SCL-independent interaction with P300/PCAF and LSD1. This allows LMO2 to accommodate, directly or indirectly, different types of interactions simultaneously, providing an extra level of regulation in complex formation.

Therefore, LMO2 appears as a versatile adaptor protein, which not only locks SCL:E47 structures and modulates their DNA-binding affinities but also, through interaction with additional regulators, plays a critical role in directing complexes to DNA. These interactions mediate additional exposure to cofactors, thereby modulating the transcriptional activities of the complex.

The SCL:LMO2 Interface: A Target for Small Inhibitory Molecules

The small size of the SCL:LMO2 interface area (620 $Å^2$) together with the presence of a defined secondary structure element (SCL H2) make it a suitable target for small inhibitory molecules (Bourgeas et al., 2010). Importantly, our mutational analyses provide a detailed picture of key residues along the SCL:LMO2 interface required for stable and synergistic interaction. Given the prevalence of human T-ALL cases coexpressing SCL and LMO2, this interface emerges as a promising target for the development of small molecules aimed at disrupting protein:protein interactions.

EXPERIMENTAL PROCEDURES

Purification, Crystallization, and Structure Determination

LMO2:LDB1-LID construct was cloned, expressed, and purified as described (El Omari et al., 2010). Histidine-tag SCL_{bHLH} and untagged $E47_{bHLH}$ were coexpressed in *E. coli* and purified by affinity and size exclusion chromatography. The complex was obtained by mixing $LMO2:LDB1_{LID}$ and $(SCL:E47)_{bHLH}$ prior to size exclusion chromatography. DNA was added as a 1.2:1 DNA-protein ratio before crystallization. Two crystal forms were obtained; the first one was solved by experimental phasing to 2.8 Å resolution and contained the DNA-bound $(SCL:E47)_{bHLH}$:$LMO2:LDB1_{LID}$. The second one, at 2.9 Å resolution, was solved by molecular replacement and contained DNA-bound $(SCL:E47)_{bHLH}$. For further details, see the Extended Experimental Procedures and Table 1.

Mammalian Two-Hybrid Assays

Plasmids encoding pVP16-SCL (VP16-transactivating domain fused to SCL) and pM-LMO2 (GAL4 DNA-binding domain fused to LMO2) have been described (El Omari et al., 2011; Schlaeger et al., 2004). E47 was fused to GAL4 DNA-binding domain. All mutants were generated with a QuikChange II Site-Directed Mutagenesis Kit (Stratagene). The CheckMate Mammalian Two-Hybrid System (Promega) was used to conduct two-hybrid experiments in HEK293 cells as described (El Omari et al., 2011).

LMO2-GFP Mutant Expression Constructs, RNA Injection, Zebrafish Embryos, and Imaging

The *gfp-Lmo2* construct (El Omari et al., 2011; Gering et al., 2003) was mutated using the QuikChange Site-Directed Mutagenesis Kit (Stratagene). Full-length *Lmo2* RNA was synthesized using the Sp6 mMessage mMachine kit (Ambion). HA-*Scl/Tal1* and *H2B-RFP* mRNA was generated as described (El Omari et al., 2011; Megason and Fraser, 2003). Zebrafish embryos were obtained from natural matings and coinjected at the one-cell stage with 100 pg of each mRNA. Embryos were raised and staged as described by Westerfield (2007). Fluorescent signal was acquired on a Zeiss Lumar V12 fluorescence microscope and in situ analysis carried out as described (El Omari et al., 2011). RNA isolation and quantitative reverse-transcription PCR are described in the Extended Experimental Procedures. The fish studies conformed to the regulatory standards approved by UK Government Home Office.

Biotin-Tagged Constructs

WT *Scl* and *Lmo2* cDNAs were tagged in their 5′ end with an oligonucleotide encoding avidin-TEV-FLAG (Driegen et al., 2005) and cloned downstream of the EF1α promoter in a vector bearing a neomycin- or zeocin-resistance marker (pEF1αbio-SCL and pEF1αbio-LMO2). Mutations in *Scl* and *Lmo2* cDNAs were generated by mutagenesis as directed by the manufacturer (QuikChange Lightning Multi Site-Directed Mutagenesis Kit; Agilent Technologies).

Cell Lines and Transfections

WT and $Scl^{-/-}$ J1 ES cells were maintained on gelatinized plates as previously described (Porcher et al., 1996). Murine erythroleukemia (MEL) cells were maintained as described by Schuh et al. (2005). Transfection procedures are described in the Extended Experimental Procedures.

Streptavidin-Affinity Pull-Downs

Nuclear extracts were prepared as described by Schuh et al. (2005). Pull-downs were performed in 50 mM Tris-HCl (pH 7.5), 0.3% NP40, 150 mM NaCl, and protease inhibitors, with 500 μg to 1 mg of nuclear extracts and 50 μl of paramagnetic streptavidin beads (Dynabeads; Invitrogen), at room temperature for 30 min (bio-SCL) or at 4°C overnight (bio-LMO2). Proteins were eluted by boiling in Laemmli buffer. Crude nuclear extracts (input [IN] 20 μg), pull-down eluate (PD), and unbound fraction ([UN] 20 μg) were analyzed by western blot.

ACCESSION NUMBERS

Atomic coordinates and structure factors for the $(SCL:E47)_{bHLH}$:LMO2:$LDB1_{LID}$:E box and the $(SCL:E47)_{bHLH}$:E box complexes have been respectively deposited in the Protein Data Bank under accession numbers 2YPA and 2YPB.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, four figures, and one table and can be found with this article online at http://dx.doi.org/10.1016/j.celrep.2013.06.008.

ACKNOWLEDGMENTS

We are grateful to the staff of beamlines I02 and I04 (Diamond Light Source) for technical support. We thank T. Ellenberger and D. Fass for providing the coordinates of the E47 homodimer bound to DNA. We are indebted to Y. Jones, D. Higgs, R. Gibbons, and T. Milne for critical reading of the manuscript. This work was funded by the Leukaemia and Lymphoma Research to E.J.M. and C.P. (08040) and to C.P. and P.V. (08068) and supported by a Wellcome Trust core award (090532/Z/09/Z) to the Wellcome Trust Centre for Human Genetics. E.J.M. is a Royal Society University Research Fellow. C.P. and R.P. are funded by the Medical Research Council (MRC). P.V. acknowledges funding from the MRC and the Department of Health's NIHR Biomedical Research Centres.

Received: February 11, 2013
Revised: April 23, 2013
Accepted: June 6, 2013
Published: July 3, 2013

REFERENCES

Aoyama, M., Ozaki, T., Inuzuka, H., Tomotsune, D., Hirato, J., Okamoto, Y., Tokita, H., Ohira, M., and Nakagawara, A. (2005). LMO3 interacts with neuronal transcription factor, HEN2, and acts as an oncogene in neuroblastoma. Cancer Res. *65*, 4587–4597.

Begley, C.G., Aplan, P.D., Denning, S.M., Haynes, B.F., Waldmann, T.A., and Kirsch, I.R. (1989). The gene SCL is expressed during early hematopoiesis and encodes a differentiation-related DNA-binding motif. Proc. Natl. Acad. Sci. USA *86*, 10128–10132.

Boehm, T., Foroni, L., Kaneko, Y., Perutz, M.F., and Rabbitts, T.H. (1991). The rhombotin family of cysteine-rich LIM-domain oncogenes: distinct members are involved in T-cell translocations to human chromosomes 11p15 and 11p13. Proc. Natl. Acad. Sci. USA *88*, 4367–4371.

Bourgeas, R., Basse, M.J., Morelli, X., and Roche, P. (2010). Atomic analysis of protein-protein interfaces with known inhibitors: the 2P2I database. PLoS One *5*, e9598.

Choi, K., Kennedy, M., Kazarov, A., Papadimitriou, J.C., and Keller, G. (1998). A common precursor for hematopoietic and endothelial cells. Development *125*, 725–732.

Condorelli, G., Vitelli, L., Valtieri, M., Marta, I., Montesoro, E., Lulli, V., Baer, R., and Peschle, C. (1995). Coordinate expression and developmental role of Id2 protein and TAL1/E2A heterodimer in erythroid progenitor differentiation. Blood *86*, 164–175.

Dastmalchi, S., Wilkinson-White, L., Kwan, A.H., Gamsjaeger, R., Mackay, J.P., and Matthews, J.M. (2012). Solution structure of a tethered Lmo2(LIM2)/Ldb1(LID) complex. Protein Sci. *21*, 1768–1774.

DeLano, W.L. (2002). The PyMOL Molecular Graphics System, Version 1.5 (Schrödinger, LLC).

Draheim, K.M., Hermance, N., Yang, Y., Arous, E., Calvo, J., and Kelliher, M.A. (2011). A DNA-binding mutant of TAL1 cooperates with LMO2 to cause T cell leukemia in mice. Oncogene *30*, 1252–1260.

Driegen, S., Ferreira, R., van Zon, A., Strouboulis, J., Jaegle, M., Grosveld, F., Philipsen, S., and Meijer, D. (2005). A generic tool for biotinylation of tagged proteins in transgenic mice. Transgenic Res. *14*, 477–482.

D'Souza, S.L., Elefanty, A.G., and Keller, G. (2005). SCL/Tal-1 is essential for hematopoietic commitment of the hemangioblast but not for its development. Blood *105*, 3862–3870.

Ellenberger, T., Fass, D., Arnaud, M., and Harrison, S.C. (1994). Crystal structure of transcription factor E47: E-box recognition by a basic region helix-loop-helix dimer. Genes Dev. *8*, 970–980.

El Omari, K., Porcher, C., and Mancini, E.J. (2010). Purification, crystallization and preliminary X-ray analysis of a fusion of the LIM domains of LMO2 and the LID domain of Ldb1. Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun. *66*, 1466–1469.

El Omari, K., Hoosdally, S.J., Tuladhar, K., Karia, D., Vyas, P., Patient, R., Porcher, C., and Mancini, E.J. (2011). Structure of the leukemia oncogene LMO2: implications for the assembly of a hematopoietic transcription factor complex. Blood *117*, 2146–2156.

Ferrando, A.A., Neuberg, D.S., Staunton, J., Loh, M.L., Huard, C., Raimondi, S.C., Behm, F.G., Pui, C.H., Downing, J.R., Gilliland, D.G., et al. (2002). Gene expression signatures define novel oncogenic pathways in T cell acute lymphoblastic leukemia. Cancer Cell *1*, 75–87.

Ferrando, A.A., Herblot, S., Palomero, T., Hansen, M., Hoang, T., Fox, E.A., and Look, A.T. (2004). Biallelic transcriptional activation of oncogenic transcription factors in T-cell acute lymphoblastic leukemia. Blood *103*, 1909–1911.

Fujiwara, T., O'Geen, H., Keles, S., Blahnik, K., Linnemann, A.K., Kang, Y.A., Choi, K., Farnham, P.J., and Bresnick, E.H. (2009). Discovering hematopoietic mechanisms through genome-wide analysis of GATA factor chromatin occupancy. Mol. Cell *36*, 667–681.

Gering, M., Yamada, Y., Rabbitts, T.H., and Patient, R.K. (2003). Lmo2 and Scl/Tal1 convert non-axial mesoderm into haemangioblasts which differentiate into endothelial cells in the absence of Gata1. Development *130*, 6187–6199.

German, M.S., Wang, J., Chadwick, R.B., and Rutter, W.J. (1992). Synergistic activation of the insulin gene by a LIM-homeo domain protein and a basic helix-loop-helix protein: building a functional insulin minienhancer complex. Genes Dev. *6*, 2165–2176.

Goodrich, J.A., and Tjian, R. (2010). Unexpected roles for core promoter recognition factors in cell-type-specific transcription and gene regulation. Nat. Rev. Genet. *11*, 549–558.

Hall, M.A., Curtis, D.J., Metcalf, D., Elefanty, A.G., Sourris, K., Robb, L., Gothert, J.R., Jane, S.M., and Begley, C.G. (2003). The critical regulator of embryonic hematopoiesis, SCL, is vital in the adult for megakaryopoiesis, erythropoiesis, and lineage choice in CFU-S12. Proc. Natl. Acad. Sci. USA *100*, 992–997.

Herblot, S., Steff, A.M., Hugo, P., Aplan, P.D., and Hoang, T. (2000). SCL and LMO1 alter thymocyte differentiation: inhibition of E2A-HEB function and pre-T alpha chain expression. Nat. Immunol. *1*, 138–144.

Kassouf, M.T., Chagraoui, H., Vyas, P., and Porcher, C. (2008). Differential use of SCL/TAL-1 DNA-binding domain in developmental hematopoiesis. Blood *112*, 1056–1067.

Kassouf, M.T., Hughes, J.R., Taylor, S., McGowan, S.J., Soneji, S., Green, A.L., Vyas, P., and Porcher, C. (2010). Genome-wide identification of TAL1's functional targets: insights into its mechanisms of action in primary erythroid cells. Genome Res. *20*, 1064–1083.

Kong, Y., Flick, M.J., Kudla, A.J., and Konieczny, S.F. (1997). Muscle LIM protein promotes myogenesis by enhancing the activity of MyoD. Mol. Cell. Biol. *17*, 4750–4760.

Lamonica, J.M., Vakoc, C.R., and Blobel, G.A. (2006). Acetylation of GATA-1 is required for chromatin occupancy. Blood *108*, 3736–3738.

Lécuyer, E., and Hoang, T. (2004). SCL: from the origin of hematopoiesis to stem cells and leukemia. Exp. Hematol. *32*, 11–24.

Lécuyer, E., Larivière, S., Sincennes, M.C., Haman, A., Lahlil, R., Todorova, M., Tremblay, M., Wilkes, B.C., and Hoang, T. (2007). Protein stability and transcription factor complex assembly determined by the SCL-LMO2 interaction. J. Biol. Chem. *282*, 33649–33658.

Lee, S.K., and Pfaff, S.L. (2003). Synchronization of neurogenesis and motor neuron specification by direct coupling of bHLH and homeodomain transcription factors. Neuron *38*, 731–745.

Li, Y., Deng, C., Hu, X., Patel, B., Fu, X., Qiu, Y., Brand, M., Zhao, K., and Huang, S. (2012). Dynamic interaction between TAL1 oncoprotein and LSD1 regulates TAL1 function in hematopoiesis and leukemogenesis. Oncogene *31*, 5007–5018.

Lo Conte, L., Chothia, C., and Janin, J. (1999). The atomic structure of protein-protein recognition sites. J. Mol. Biol. *285*, 2177–2198.

Longo, A., Guanga, G.P., and Rose, R.B. (2008). Crystal structure of E47-NeuroD1/beta2 bHLH domain-DNA complex: heterodimer selectivity and DNA recognition. Biochemistry *47*, 218–229.

Luscombe, N.M., Laskowski, R.A., and Thornton, J.M. (1997). NUCPLOT: a program to generate schematic diagrams of protein-nucleic acid interactions. Nucleic Acids Res. *25*, 4940–4945.

Ma, P.C., Rould, M.A., Weintraub, H., and Pabo, C.O. (1994). Crystal structure of MyoD bHLH domain-DNA complex: perspectives on DNA recognition and implications for transcriptional activation. Cell *77*, 451–459.

Malik, S., and Roeder, R.G. (2010). The metazoan Mediator co-activator complex as an integrative hub for transcriptional regulation. Nat. Rev. Genet. *11*, 761–772.

Manetopoulos, C., Hansson, A., Karlsson, J., Jönsson, J.I., and Axelson, H. (2003). The LIM-only protein LMO4 modulates the transcriptional activity of HEN1. Biochem. Biophys. Res. Commun. *307*, 891–899.

Massari, M.E., and Murre, C. (2000). Helix-loop-helix proteins: regulators of transcription in eucaryotic organisms. Mol. Cell. Biol. *20*, 429–440.

Megason, S.G., and Fraser, S.E. (2003). Digitizing life at the level of the cell: high-performance laser-scanning microscopy and image analysis for in toto imaging of development. Mech. Dev. *120*, 1407–1420.

Palii, C.G., Perez-Iratxeta, C., Yao, Z., Cao, Y., Dai, F., Davison, J., Atkins, H., Allan, D., Dilworth, F.J., Gentleman, R., et al. (2011). Differential genomic targeting of the transcription factor TAL1 in alternate haematopoietic lineages. EMBO J. *30*, 494–509.

Patterson, L.J., Gering, M., Eckfeldt, C.E., Green, A.R., Verfaillie, C.M., Ekker, S.C., and Patient, R. (2007). The transcription factors Scl and Lmo2 act together during development of the hemangioblast in zebrafish. Blood *109*, 2389–2398.

Plevin, M.J., Zhang, J., Guo, C., Roeder, R.G., and Ikura, M. (2006). The acute myeloid leukemia fusion protein AML1-ETO targets E proteins via a paired amphipathic helix-like TBP-associated factor homology domain. Proc. Natl. Acad. Sci. USA *103*, 10242–10247.

Porcher, C., Swat, W., Rockwell, K., Fujiwara, Y., Alt, F.W., and Orkin, S.H. (1996). The T cell leukemia oncoprotein SCL/tal-1 is essential for development of all hematopoietic lineages. Cell *86*, 47–57.

Porcher, C., Liao, E.C., Fujiwara, Y., Zon, L.I., and Orkin, S.H. (1999). Specification of hematopoietic and vascular development by the bHLH transcription factor SCL without direct DNA binding. Development *126*, 4603–4615.

Ryan, D.P., Duncan, J.L., Lee, C., Kuchel, P.W., and Matthews, J.M. (2008). Assembly of the oncogenic DNA-binding complex LMO2-Ldb1-TAL1-E12. Proteins *70*, 1461–1474.

Sanda, T., Lawton, L.N., Barrasa, M.I., Fan, Z.P., Kohlhammer, H., Gutierrez, A., Ma, W., Tatarek, J., Ahn, Y., Kelliher, M.A., et al. (2012). Core transcriptional regulatory circuit controlled by the TAL1 complex in human T cell acute lymphoblastic leukemia. Cancer Cell *22*, 209–221.

Schlaeger, T.M., Schuh, A., Flitter, S., Fisher, A., Mikkola, H., Orkin, S.H., Vyas, P., and Porcher, C. (2004). Decoding hematopoietic specificity in the helix-loop-helix domain of the transcription factor SCL/Tal-1. Mol. Cell. Biol. *24*, 7491–7502.

Schoenebeck, J.J., Keegan, B.R., and Yelon, D. (2007). Vessel and blood specification override cardiac potential in anterior mesoderm. Dev. Cell *13*, 254–267.

Schuh, A.H., Tipping, A.J., Clark, A.J., Hamlett, I., Guyot, B., Iborra, F.J., Rodriguez, P., Strouboulis, J., Enver, T., Vyas, P., and Porcher, C. (2005). ETO-2 associates with SCL in erythroid cells and megakaryocytes and provides repressor functions in erythropoiesis. Mol. Cell. Biol. *25*, 10235–10250.

Shirakata, M., Friedman, F.K., Wei, Q., and Paterson, B.M. (1993). Dimerization specificity of myogenic helix-loop-helix DNA-binding factors directed by nonconserved hydrophilic residues. Genes Dev. *7*(12A), 2456–2470.

Simões, F.C., Peterkin, T., and Patient, R. (2011). Fgf differentially controls cross-antagonism between cardiac and haemangioblast regulators. Development *138*, 3235–3245.

Snow, J.W., and Orkin, S.H. (2009). Translational isoforms of FOG1 regulate GATA1-interacting complexes. J. Biol. Chem. *284*, 29310–29319.

Soler, E., Andrieu-Soler, C., de Boer, E., Bryne, J.C., Thongjuea, S., Stadhouders, R., Palstra, R.J., Stevens, M., Kockx, C., van Ijcken, W., et al. (2010). The genome-wide dynamics of the binding of Ldb1 complexes during erythroid differentiation. Genes Dev. *24*, 277–289.

Tripic, T., Deng, W., Cheng, Y., Zhang, Y., Vakoc, C.R., Gregory, G.D., Hardison, R.C., and Blobel, G.A. (2009). SCL and associated proteins distinguish active from repressive GATA transcription factor complexes. Blood *113*, 2191–2201.

Van Handel, B., Montel-Hagen, A., Sasidharan, R., Nakano, H., Ferrari, R., Boogerd, C.J., Schredelseker, J., Wang, Y., Hunter, S., Org, T., et al. (2012). Scl represses cardiomyogenesis in prospective hemogenic endothelium and endocardium. Cell *150*, 590–605.

Wadman, I., Li, J., Bash, R.O., Forster, A., Osada, H., Rabbitts, T.H., and Baer, R. (1994). Specific in vivo association between the bHLH and LIM proteins implicated in human T cell leukemia. EMBO J. *13*, 4831–4839.

Wadman, I.A., Osada, H., Grütz, G.G., Agulnick, A.D., Westphal, H., Forster, A., and Rabbitts, T.H. (1997). The LIM-only protein Lmo2 is a bridging molecule assembling an erythroid, DNA-binding complex which includes the TAL1, E47, GATA-1 and Ldb1/NLI proteins. EMBO J. *16*, 3145–3157.

Westerfield, M. (2007). The Zebrafish Book. A Guide for the Laboratory Use of Zebrafish (*Danio rerio*), Fifth Edition (Eugene, OR: University of Oregon Press).

Zhang, J., Kalkum, M., Yamamura, S., Chait, B.T., and Roeder, R.G. (2004). E protein silencing by the leukemogenic AML1-ETO fusion protein. Science *305*, 1286–1289.